DOCUMENT RESUME

ED 063 845 FL 003 384

AUTHOR Natalicio, Diana S.; Williams, Frederick
TITLE Oral Language Assessment.
PUB DATE 6 Apr 72
NOTE 14p.; Paper presented at the annual meeting of the American Educational Research Association, April 6, 1972, Chicago, Illinois

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Child Language; *Children; Dialect Studies; Early Childhood; Language; Language Research; Linguistic Competence; *Linguistic Performance; Mexican Americans; *Minority Groups; Negroes; Preschool Children; Speech; Standard Spoken Usage; Tables (Data); *Teacher Education; *Test Reliability

ABSTRACT

This paper reports the attempt to see which characteristics of the speech of Black and Mexican American children would be reliably evaluated by experts specializing in dialect study. Presumably, if selected characteristics were evaluated with consistency and bases for these evaluations were given, such results could serve in training teachers to recognize and deal with language difference in minority group children. Evaluations for both language groups were in terms of judgments concerning language dominance and Standard American English comprehension, production, phonology, intonation, inflection, syntax, possible language pathologies, and predictions of reading achievement. In addition, the Mexican American children were evaluated on Spanish comprehension, production, phonology, intonation, and syntax. Reliability estimates are provided for each of the aspects of the investigation. (Author/VM)

ED 063845

ORAL LANGUAGE ASSESSMENT

Diana S. Natalicio
University of Texas at El Paso

and

Frederick Williams
University of Texas at Austin

Paper presented to symposium, "Oral Language Differences: Linguistic and Attitudinal Correlates" (Program C-72, Session 19.5)

1972 AERA Convention

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

FL003 384

ABSTRACT. This paper reports the attempt to assess which characteristics of the speech of Black and Mexican-American children (grades K-2) could be reliably evaluated by experts specializing in dialect study. Tapes of ten Black and ten Mexican-American children who had responded to a set of commercially available test materials were evaluated by the experts. Evaluations for both groups were in terms of judgments (scale ratings) of language dominance, comprehension, production, phonology, intonation, inflectional endings, syntax, language pathologies, and predictions of reading achievement. For each scaled evaluation, evaluators provided a description of their bases for judgment. Results indicated high reliability of scale judgments except for ratings of intonation, language pathologies and for predictions of reading achievement. The comments which served as bases for making scale judgments were highly consistent with language differences typically identified in the two linguistic communities represented, and were congruent with the scale ratings themselves. The results are interpreted in terms of their application to training teachers to recognize and deal effectively with language differences in minority group children.

In the 1960s linguists, psychologists, and educators acknowledged the importance of focusing on the language competencies of children entering the educational system for the first time. Indeed, oral language seems to be the single most important aspect of such diverse efforts as Head Start and Sesame Street, designed for the preschool child. The target of such special programs has been the "atypical" child whose socioeconomic status or ethnic background differs from that of the "average" child for whom most educational curricula have been designed.

Two distinct schools of thought arose out of a common concern for "atypical" children. The first, as perhaps best exemplified in the work of Deutsch (1967) and Bereiter and Englemann (1966), views the "atypical" child as having a language <u>deficit</u> which must be made up if the child is to have an equal opportunity in the average classroom; the obvious solution for a proponent of this position is the design and implementation of compensatory programs such as Head Start which will provide children with the means to make up the deficit before entering the regular educational process. Proponents of the <u>difference</u> position are, of course, opposed to any notion of deficiency, holding that "atypical" children are different in many respects, including language, and that it is up to the educational system to deal with these differences rather than to attempt to force the child to compensate for his background. This position is exemplified in the writings of Baratz (1970) and Labov (1970).

What is interesting and even disturbing about such debates is that they so seldom result in a change in classroom teacher behavior. Thus, although there appears to be a growing acceptance of the <u>difference</u> position among linguists and psychologists, and although classroom teachers

may be aware of this trend, they are often ill-equipped to bring about the innovations in their teaching strategies which would reflect this general orientation.

This paper reports the attempt to see which characteristics of the speech of Black and Mexican-American children would be reliably evaluated by experts specializing in dialect study. Also there was the attempt to have the experts report the bases of their evaluations. Presumably, if selected characteristics were evaluated with consistency, and bases for these evaluations were given, such results could serve in training teachers to recognize and deal with language differences in minority group children.

Procedures

Oral language performances on a set of commercially available sentence repetition test materials[1] recorded on tape by children in grades K-2 in San Antonio, Texas were reviewed, and the performances of ten Black and ten Mexican-American children were selected to represent the entire corpus of 750 recordings. Experts, defined as persons whose professional activities showed evidence of interest and expertise in the areas of child language and social dialects, were contacted as potential evaluators of the recorded performances. Fifteen persons evaluated the 10 Black language samples, and fourteen evaluated the 10 Mexican-American language samples. Evaluations for both language groups were in terms of

---

[1] From Gloria & David Beginning English Series No. 20, 1958; Gloria & David Beginning Spanish Series No. 40, 1959. Copyright © Language Arts, Inc. These materials and the instrumentation (a sound and picture synchronized cartridge and a receiver unit) used to administer them were selected on the basis of the facility with which sentence imitation data may be elicited.

judgments concerning language dominance, SAE (Standard American English) comprehension, SAE production, SAE phonology, SAE intonation, SAE inflections, SAE syntax, possible language pathologies, and predictions of reading achievement. In addition, the Mexican-American children were evaluated on Spanish comprehension, Spanish production, Spanish phonology, Spanish intonation, and Spanish syntax. A seven-point scale was provided the evaluators for their judgments on each of the above areas in each child's performance. For each scaled evaluation the experts provided a description of the aspects of each performance which served as bases for judgment on each of the scales and the utterances in the sentence-repetition task which exemplified a given aspect of performance. For example, a questionnaire item submitted to the experts took the following form:

A. How would you rate this child's overall mastery of (e.g., comprehension of SAE)

Good ____: ____: ____: ____: ____: ____: ____ Bad

B. Upon which aspects of this child's performance did you base your rating? Please be specific.

Aspect	As in:	Aspect	As in:
_____	_____	_____	_____
_____	_____	_____	_____

Results: Evaluations of Reliability

By assigning numbers to the scaled ratings, it was then possible to calculate a mathematical index of reliability (Ebel, 1951; Veldman, 1970) which would vary between 0.0 (no reliability) and 1.0 (perfect reliability). For practical interpretation here, an index of from .90 to 1.0 was interpreted as of high reliability; .80 to .89 of moderate reliability, and anything lower of questionable or low reliability.

Insert Table 1

Table 1 summarizes the reliability results for the 10 items referring to the evaluations of the Black children's samples. The scales showing the highest reliability are those relative to dominance of SAE and Black dialect (.95 and .94, respectively). These are closely followed by the SAE inflection and production scales (.92). The reliabilities of ratings on phonology, syntax, and overall comprehension of SAE were moderate, all exceeding .85. The three ratings showing questionable reliability are those relative to pathologies, intonation, and prediction of reading achievement.

The estimated reliability of ratings provided by the fourteen evaluators for the ten language samples from Mexican-American children appear in Table 2.

Insert Table 2

The highest reliability estimates for the ratings of the ten Mexican-American language samples obtain in the areas of Spanish dominance, Spanish syntax, SAE comprehension, SAE inflections, Spanish comprehension, Spanish production, SAE syntax, and SAE production; all of these estimates of reliability fall within the high range. As in the case of the Black language samples, the three areas for which estimated reliability of ratings was low were for SAE intonation, pathologies, and reading predictions.

In examining these reliability estimates, it should be emphasized that they represent the consistencies obtaining in the ratings provided for each child with respect to each of the linguistic aspects (questionnaire

items) included in this study. The high reliability estimates obtained here indicate great consistency in the ratings of the same child's performance by fourteen or fifteen different evaluators. The recorded performances elicited by this sentence repetition task thus do seem to permit independent evaluations with a high degree of reliability. These aspects of performance are good topics for teacher training in evaluation of the sentence imitations.

Results: Bases for Evaluations

Considerable consistency was also observed in those aspects of each child's performance cited by the experts as bases for assigning ratings to performances. Specific aspects of performance cited by the evaluator panels will be divided into two categories, phonology and grammar, and will be presented separately for each of the two samples, Black and Mexican-American.

Those aspects cited as relevant to the evaluation of Black children's performances which demonstrated a high level of consistency among the fifteen evaluators included:

Grammar:

1. Deletion of inflectional ending indicating the third person present tense of verbs ("goes" produced as "go," "helps" as "help").
2. Substitution of subject pronoun for possessive pronoun ("she head" for "her head"). In addition, it was frequently observed that the substitution of possessive pronouns involved gender undifferentiation where the subject pronoun used in place of the possessive violated the concord with the gender

of the subject pronoun of the sentence ("She has soap on he head.").

3. Replacement of third person singular form /haez/ by [haev] or [haef].
4. Deletion of the noun possessive marker in pre-noun position ("David's neck" replaced by "David neck").
5. Deletion of "is" and "are" as part of auxiliary ("is going" replaced by "going"). "Is" used with plural subject. "Ain't" replaced "is not."
6. Deletion of the noun plural marker ("shoes" replaced by "shoe"). Use of hyper-plurals ("feets," "teeths").
7. Substitution for subject pronouns ("Her has the soap.").

Phonology:

1. /ɖ/ replaced by /d/, especially in initial position ([dey] for "they").
2. /θ/ replaced by /f/ or /s/ or /t/ ([tiys] for "teeth").
3. /ɛ/ as in "bed" lengthened and diphthongized ([beyd] for "bed").
4. /l/ and /r/ interchanged, particularly when occurring as the second member of a consonant cluster ([krowz] for "clothes").
5. Consonant clusters, both initial and final, reduced to a single consonant ([kuwl] for "school" and [liyn] for "clean").
6. Final voiced stops devoiced ([bɛt] for "bed").
7. Final voiceless stops deleted ([lay] for "light").
8. Mid-central vowel /ə/ fronted to /ɛ/ ([brɛš̌] for "brush").

Aspects of Mexican-American children's performances cited with consistency as relevant to overall performances by the experts included the following:

Grammar:

1. Deletion of inflectional ending indicating the third person, present tense of verbs ("goes" produced as "go"; "helps" as "help").
2. Deletion of the noun plural marker ("shoes" replaced by "shoe"). Use of hyper-plurals ("feets," "teeths").
3. Deletion of the noun possessive marker in pre-noun position ("David's neck" replaced by "David neck").
4. Substitution of either subject pronoun or article for possessive pronoun ("she head" or "the head" for "her head").
5. Replacement of third person singular form /haez/ by [haev] or [haef].

Phonology:

1. Substitution of /č/ for /š/ ("washes" replaced by "watches").
2. Initial /ð/ replaced by /d/ ([dey] for "they"). Intervocalic /ð/ (as in "mother") weakened so as to resemble a vowel glide.
3. Replacement of voiced /z/ by /s/ ([s̬uws] for "shoes").
4. Reduction of initial and final consonant clusters ([kuwl] for "school").
5. Substitution of [f] and [s] for /θ/ ([tiyf] for "teeth").
6. No differentiation among low and central vowels, /ae/, /ə/, /a/, and /ɔ/ ([brač] for "brush").
7. Unaspirated voiceless stops in initial position.

8. No differentiation between /i/ and /iy/ (as in "fit" and "feet," respectively).
9. Vowels and vowel glides reduced in length.
10. Final voiced stops devoiced.

An examination of the specific performance aspects cited by a majority of the evaluators rating each of the two language groups shows nonstandard features shared by the two language groups, especially in the area of grammar, as well as features which differ between them. For example, both Black and Mexican-American children's performances were reported to reflect the deletion of various inflectional endings (the third person present tense of verbs, noun plurals, and noun possessives) and some confusion over possessive and subject pronouns. Certain common features were also shared by both groups on the phonological level, e.g., replacing /d/ and /θ/, and the reduction of consonant clusters. However, there were some significant differences between the two language samples on this level. Among these differences were that Black children were reported to lengthen normally short vowels and even to diphthongize them, and Mexican-American children were reported to shorten normally long vowels and reduce diphthongs to a single short vowel sound. Black children were also reported to front the mid-central vowel /ə/ to /ɛ/, and Mexican-American children lowered this same vowel, /ə/, to /a/, resulting in the Black child's rendition of "brush" sounding like [brɛs̆] and the Mexican-American child's like [bras̆] or [brac̆].

Discussion

The high consistency in the ratings assigned to given aspects of each child's performance by the two evaluator panels provides a basis

for determining which aspects of language teachers might be trained to evaluate. Those aspects for which low reliability estimates were obtained from the evaluator panels should probably be avoided in teacher preparation programs because even the expert panels were unable to arrive at a consensus on them. The fact that evaluator panels agreed on not only the specific ratings which they assigned to most aspects of given performances, but also on the performance features upon which those ratings were assigned, indicates that a training program focusing on the experts' criteria should achieve a high level of reliability among teacher trainees.

It should be borne in mind that these evaluations were based upon a fixed set of sentences drawn from a commercially available test package. Thus, it may be that if further sentences or test items were incorporated, some types of evaluation might be added or some of the evaluations reported here might improve in reliability. On the other hand, the present results do provide a basis for direct application in teacher training. Using sentence imitation examples from the present research, teacher trainees can observe the children's responses along with the experts' evaluations. By being informed of the bases of experts' evaluations, teachers should be able to gain some practical degree of familiarity with the special characteristics of the speech of linguistically different children and be able to evaluate such characteristics. Teacher ability in this task can itself be evaluated by comparing a teacher's evaluations with those supplied by the experts.

REFERENCES

Baratz, J. C. Teaching reading in an urban negro school system. In LANGUAGE AND POVERTY, F. Williams (Ed.). Chicago: Markham Publishing Company, 1970, pp. 11-24.

Bereiter, C. and Englemann, S. TEACHING DISADVANTAGED CHILDREN IN THE PRESCHOOL. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Deutsch, M., and associates. THE DISADVANTAGED CHILD. New York: Basic Books, 1967.

Ebel, R. L. Estimation of the reliability of ratings, PSYCHOMETRIKA, 16, 4, 1951, pp. 407-424.

Labov, W. The logic of nonstandard English. In LANGUAGE AND POVERTY, F. Williams (Ed.). Chicago: Markham Publishing Company, 1970, pp. 153-189.

Natalicio, D. and Williams, F. REPETITION AS AN ORAL LANGUAGE ASSESSMENT TECHNIQUE. Austin, Texas: Center for Communication Research, 1970. Available as ERIC document ED051680.

Veldman, D. EDSTAT-V, Basic statistical computer programs for the CDC6600, University of Texas at Austin, 1970 (mimeo).

Williams, F. LANGUAGE AND POVERTY. Chicago: Markham Publishing Company, 1970.

Table 1 Reliability estimates based on ratings of fifteen evaluators of Black language sample

Aspect of performance	Average reliability estimate (15 raters)
1. Black dialect dominance (strong-weak)	.94
2. SAE dominance (strong-weak)	.95
3. SAE comprehension (good-bad)	.86
4. SAE production (good-bad)	.92
5. Pathologies (Yes-No)	.69
6. SAE phonology (good-bad)	.88
7. SAE intonation (good-bad)	.55
8. SAE inflections (good-bad)	.92
9. SAE syntax (good-bad)	.87
10. Predict reading achievement (Yes-No)	.47

Table 2 Reliability estimates based on ratings of fourteen evaluators of Mexican-American language sample

Aspect of performance	Average reliability estimate (14 raters)
1. Spanish dominance (strong-weak)	.96
2. SAE dominance (strong-weak)	.93
3. SAE comprehension (good-bad)	.95
4. Spanish comprehension (good-bad)	.95
5. SAE production (good-bad)	.94
6. Spanish production (good-bad)	.95
7. Pathologies (Yes-No)	.19
8. SAE phonology (good-bad)	.91
9. Spanish phonology (good-bad)	.93
10. SAE intonation (good-bad)	.78
11. Spanish intonation (good-bad)	.90
12. SAE inflections (good-bad)	.95
13. SAE syntax (good-bad)	.94
14. Spanish syntax (good-bad)	.95
15. Predict reading achievement (Yes-No)	0.00